

Compact Name

Evaluator Name

Evaluation Title

Study Documentation

(Note: MCC may make slight revisions to this template over time)

Date Created

Metadata Production

Metadata Producer(s)	Millennium Challenge Corporation (MCC) , Review of Metadata
Production Date	Date Created
Version	Version 1.0 (Date Created)
Identification	Ex. DDI-MCC-BFA-A2F-RuralFinance-2015-v1

Table of Contents

Overview	4
Scope & Coverage	4
Producers & Sponsors	4
Sampling	4
Data Collection	5
Data Processing & Appraisal	7
Accessibility	8
Files Description	10
Variables List	11

Evaluation Title*Translated Evaluation Title***Overview****Type** Impact or performance evaluation?**Identification** Ex. DDI-MCC-BFA-A2F-RuralFinance-2015-v1**Version** Raw data for internal use only**Program**

Brief description of the Compact components being evaluated (Project/Activity/Sub-Activity). This is essentially the first paragraph in the MCC Summary of Findings.

Program Logic

Overview of Projects/Activities/Sub-Activities objectives and a description of the causal chain from inputs, outputs, to outcomes in order to achieve objectives.

Program Participants

Please define the program participant in terms of specific selection criteria for participation in the program.

Example:

The survey covered all de jure household members (usual residents), all women aged 15-49 years resident in the household, and all children aged 0-4 years (under age 5) resident in the household.

Evaluation Summary

Overview of Evaluation Design, includes evaluation type (performance/impact), evaluation questions, exposure to treatment. Upon completion of evaluation, include brief summary of results.

Evaluation Methodology

Evaluation methodology (ex. Randomization)

Evaluation Summary**Kind of Data** Kind of data (ex. sample survey data)**Unit of Analysis** Unit of analysis. Choose from any of the following: individuals, households, community, administrative units, enterprise, school, health center, other.**Scope & Coverage**

Countries	Burkina Faso
<u>Geographic Coverage</u> This field aims at describing at what geographic level the data are representative. Typical entries will be "National coverage", "Urban (or rural) areas only", "State of ...", "Capital city", etc. Note that we do not describe here where the data was collected. For example, as sample survey could be declared as "national coverage" even in cases where some districts were not included in the sample, as long as the sampling strategy was such that the representativity is national. MCC Clarification: This description should indicate where the program is being implemented (nationally, regionally, etc) with specifically defined regions.	
<u>Universe</u> Define study population. Example: Parents of and children 0-5 years old living in rural Tanzania.	

Producers & Sponsors	
Primary Investigator(s)	Evaluator Name
Funding Agency/ies	Millennium Challenge Corporation (MCC)

Sampling
<u>Sampling Procedure</u> This field only applies to sample surveys. Information on sampling procedure is crucial (although not applicable for censuses and administrative datasets). Content and Format: Description of power calculations, sample sizes, selection process, level of representation, etc.

MCC Clarification:

Information on sampling procedure is crucial (although not applicable for censuses and administrative datasets). This section should include summary information that includes though is not limited to:

Required sample size and actual sample size,
 Selection process (e.g., probability proportional to size or over sampling),
 Stratification (implicit and explicit),
 Stages of sample selection,
 Level of representation,
 Strategy for absent respondents/not found/refusals (replacement or not),
 Sample frame used, and listing exercise conducted to update it

Example:

5000 households were selected for the sample. Of these, 4996 were occupied households and 4811 were successfully interviewed for a response rate of 96.3%. Within these households, 7815 eligible women aged 15-49 were identified for interview, of which 7505 were successfully interviewed (response rate 96.0%), and 3242 children aged 0-4 were identified for whom the mother or caretaker was successfully interviewed for 3167 children (response rate 97.7%). These give overall response rates (household response rate times individual response rate) for the women's interview of 92.5% and for the children's interview of 94.1%

Deviations from Sample Design

This field only applies to sample surveys.

MCC Clarification:

Sometimes the reality of the field requires a deviation from the sampling design (for example due to difficulty to access to zones due to weather problems, political instability, etc). If for any reason, the sample design has deviated, this should be reported here.

Response Rate**MCC Clarification:**

Response rate provides that percentage of households (or other sample unit) that participated in the survey based on the original sample size. Omissions may occur due to refusal to participate, impossibility to locate the respondent, or other. Sometimes, a household may be replaced by another by design. Check that the information provided here is consistent with the sample size indicated in the "Sampling procedure" field and the number of records found in the dataset (for example, if the sample design mention a sample of 5,000 households and the data on contain data on 4,500 households, the response rate should not be 100 percent).

Weighting

This field only applies to sample surveys.

MCC Clarification:

Provide here the list of variables used as weighting coefficient. If more than one variable is a weighting variable, describe how these variables differ from each other and what the purpose of each one of them is.

Example:

Sample weights for the household data were computed as the inverse of the probability of selection of the household, computed at the sampling domain level (urban/rural within each region). The household weights were adjusted for non-response at the domain level, and were then normalized by a constant factor so that the total weighted number of households equals the total unweighted number of households. The household weight variable is called HHWEIGHT and is used with the HH data and the HL data.

Data Collection**Data Collection Notes****MCC Clarification:**

Include any additional information and direct users to Data Collection Reports if available.

Example:

The pre-test for the survey took place from August 15, 2006 - August 25, 2006 and included 14 interviewers who would later become supervisors for the main survey.

Each interviewing team comprised of 3-4 female interviewers (no male interviewers were used due to the sensitivity of the subject matter), together with a field editor and a supervisor and a driver. A total of 52 interviewers, 14 supervisors and 14 field editors were used. Data collection took place over a period of about 6 weeks from September 2, 2006 until October 17, 2006. Interviewing took place everyday throughout the fieldwork period, although interviewing teams were permitted to take one day off per week.

Interviews averaged 35 minutes for the household questionnaire (excluding salt testing), 23 minutes for the women's questionnaire, and 27 for the under five children's questionnaire (excluding the anthropometry). Interviews were conducted primarily in English and Mumbo-jumbo, but occasionally used local translation in double-Dutch, when the respondent did not speak English or Mumbo-jumbo.

Six staff members of GenCenStat provided overall fieldwork coordination and supervision. The overall field coordinator was Mrs. Doe.

Data Collection Dates

Content and Format:

YYYY-MM-DD (Baseline); YYYY-MM-DD (Interim); YYYY-MM-DD (End-line)

Questionnaires

List of questionnaires and short description of each (languages, modules)

Example:

The questionnaires for the Generic MICS were structured questionnaires based on the MICS3 Model Questionnaire with some modifications and additions. A household questionnaire was administered in each household, which collected various information on household members including sex, age, relationship, and orphanhood status. The household questionnaire includes household characteristics, support to orphaned and vulnerable children, education, child labour, water and sanitation, household use of insecticide treated mosquito nets, and salt iodization, with optional modules for child discipline, child disability, maternal mortality and security of tenure and durability of housing.

In addition to a household questionnaire, questionnaires were administered in each household for women age 15-49 and children under age five. For children, the questionnaire was administered to the mother or caretaker of the child.

The women's questionnaire include women's characteristics, child mortality, tetanus toxoid, maternal and newborn health, marriage, polygyny, female genital cutting, contraception, and HIV/AIDS knowledge, with optional modules for unmet need, domestic violence, and sexual behavior.

The children's questionnaire includes children's characteristics, birth registration and early learning, vitamin A, breastfeeding, care of illness, malaria, immunization, and anthropometry, with an optional module for child development.

The questionnaires were developed in English from the MICS3 Model Questionnaires, and were translated into Mumbo-jumbo. After an initial review the questionnaires were translated back into English by an independent translator with no prior knowledge of the survey. The back translation from the Mumbo-jumbo version was independently reviewed and compared to the English original. Differences in translation were reviewed and resolved in collaboration with the original translators. The English and Mumbo-jumbo questionnaires were both piloted as part of the survey pretest.

All questionnaires and modules are provided as external resources.

Data Collector(s) Data Collection firm name (Abbreviation) , Affiliation

Supervision

MCC Clarification:

Please describe how data collection teams were organized (how many team members per team, how many teams per supervisor, how was data entry managed)

Example:

Interviewing was conducted by teams of interviewers. Each interviewing team comprised of 3-4 female interviewers, a field editor and a supervisor, and a driver. Each team used a 4 wheel drive vehicle to travel from cluster to cluster (and where necessary within cluster).

The role of the supervisor was to coordinate field data collection activities, including management of the field teams, supplies and equipment, finances, maps and listings, coordinate with local authorities concerning the survey plan and make arrangements for accommodation and travel. Additionally, the field supervisor assigned the work to the interviewers, spot checked work, maintained field control documents, and sent completed questionnaires and progress reports to the central office.

The field editor was responsible for reviewing each questionnaire at the end of the day, checking for missed questions, skip errors, fields incorrectly completed, and checking for inconsistencies in the data. The field editor also observed interviews and conducted review sessions with interviewers.

Responsibilities of the supervisors and field editors are described in the Instructions for Supervisors and Field Editors, together with the different field controls that were in place to control the quality of the fieldwork.

Field visits were also made by a team of central staff on a periodic basis during fieldwork. The senior staff of GenCenStat also made 3 visits to field teams to provide support and to review progress.

Data Processing & Appraisal

Data Editing

MCC Clarification:

Evaluator please describe the data cleaning process. E.g., consistency checking, wildcode checking, etc. Include the role of the data collection firm (if any) and the process used to produce the cleaned, analysis file.

The data editing should contain information on how the data was treated or controlled for in terms of consistency and coherence. This item does not concern the data entry phase but only the editing of data whether manual or automatic.

- Was a hot deck or a cold deck technique used to edit the data?
- Were corrections made automatically (by program), or by visual control of the questionnaire?
- What software was used?

Example:

Data editing took place at a number of stages throughout the processing, including:

- a) Office editing and coding
- b) During data entry
- c) Structure checking and completeness
- d) Secondary editing
- e) Structural checking of SPSS data files

Detailed documentation of the editing of data can be found in the "Data processing guidelines" document provided as an external resource.

Other Processing

MCC Clarification:

Use this field to provide as much information as possible on the data entry design. This includes such details as:

- Mode of data entry (manual or by scanning, in the field/in regions/at headquarters)
- Computer architecture (laptop computers in the field, desktop computers, scanners, PDA, other; indicate the number of computers used)
- Software used

- Use (and rate) of double data entry
- Average productivity of data entry operators; number of data entry operators involved and their work schedule

Example:

Data entry was conducted by 12 data entry operators in tow shifts, supervised by 2 data entry supervisors, using a total of 7 computers (6 data entry computers plus one supervisors' computer). All data entry was conducted at the GenCenStat head office using manual data entry. For data entry, CSPro version 2.6.007 was used with a highly structured data entry program, using system controlled approach that controlled entry of each variable. All range checks and skips were controlled by the program and operators could not override these. A limited set of consistency checks were also included in the data entry program. In addition, the calculation of anthropometric Z-scores was also included in the data entry programs for use during analysis. Open-ended responses ("Other" answers) were not entered or coded, except in rare circumstances where the response matched an existing code in the questionnaire.

100% verification of all variables was performed using independent verification, i.e. double entry of data, with separate comparison of data followed by modification of one or both datasets to correct keying errors by original operators who first keyed the files. score and quintiles were included in the datasets for use in tabulations.

Estimates of Sampling Error

MCC Clarification:

For sampling surveys, it is good practice to calculate and publish sampling error. This field is used to provide information on these calculations. This includes:

- A list of ratios/indicators for which sampling errors were computed.
- Details regarding the software used for computing the sampling error, and reference to the programs used (to be provided as external resources) as the program used to perform the calculations.
- Reference to the reports or other document where the results can be found (to be provided as external resources).

Example:

Estimates from a sample survey are affected by two types of errors: 1) non-sampling errors and 2) sampling errors. Non-sampling errors are the results of mistakes made in the implementation of data collection and data processing. Numerous efforts were made during implementation of the 2005-2006 MICS to minimize this type of error, however, non-sampling errors are impossible to avoid and difficult to evaluate statistically.

If the sample of respondents had been a simple random sample, it would have been possible to use straightforward formulae for calculating sampling errors. However, the 2005-2006 MICS sample is the result of a multi-stage stratified design, and consequently needs to use more complex formulae. The SPSS complex samples module has been used to calculate sampling errors for the 2005-2006 MICS. This module uses the Taylor linearization method of variance estimation for survey estimates that are means or proportions. This method is documented in the SPSS file CSDescriptives.pdf found under the Help, Algorithms options in SPSS.

Sampling errors have been calculated for a select set of statistics (all of which are proportions due to the limitations of the Taylor linearization method) for the national sample, urban and rural areas, and for each of the five regions. For each statistic, the estimate, its standard error, the coefficient of variation (or relative error -- the ratio between the standard error and the estimate), the design effect, and the square root design effect (DEFT -- the ratio between the standard error using the given sample design and the standard error that would result if a simple random sample had been used), as well as the 95 percent confidence intervals (+/-2 standard errors).

Accessibility

Contact(s) Monitoring & Evaluation Division of the Millennium Challenge Corporation , impact-eval@mcc.gov

Confidentiality

Citation Requirements

Citation requirement is the way that the dataset should be referenced when cited in any publication. Every dataset should have a citation requirement. This will guarantee that the data producer gets proper credit, and that analytical results can be

linked to the proper version of the dataset. The Access Policy should explicitly mention the obligation to comply with the citation requirement. The citation should include at least the primary investigator, the name and abbreviation of the dataset, the reference year, and the version number. Include also a website where the data or information on the data is made available by the official data depositor.

Content and Format:

Use of the dataset must be acknowledged using a citation which would include:

- the Identification of the Primary Investigator
- the title of the survey (including country, acronym and year of implementation)
- the survey reference number
- the source and date of download

Example:

Chicago (16th edition) (author-date)

Smith, Tom W., Peter V. Marsden, and Michael Hout. 2011. General Social Survey, 1972-2010 Cumulative File. ICPSR31521-v1. Chicago, IL: National Opinion Research Center. Distributed by Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR31521.v1

Files Description

Dataset contains 0 file(s)

Variables List

Dataset contains 0 variable(s)